Supplemental Materials

To advance the trustworthiness of the study and findings, we offer supplemental *thick description* of our methods and approaches, and report additional analyses not presented in the main manuscript for lack of space. Thick description involves clear explanations of the research process, including data collection, analysis, and reporting (Li, 2004; Stahl & King, 2020). These supplemental materials address, in detail, the following components of trustworthiness: **confirmability**, **credibility**, **dependability**, **transferability**, **triangulation**, and **negative case** comparisons.

Supplemental Methods

Community Advisory Board

The project, which was funded by the National Institutes of Mental Health (R01 MH111650 - 01 A1), was overseen by a Community Advisory Board (CAB). The CAB included leaders of organizations providing services to persons with mental illness and/or their employers; workers with lived experience of serious mental illness; and members of the academy with expertise in disability and employment. The role of the CAB was to provide peer input and debriefing for the entire research process, including: development of research questions, identification of target population, formation of the sampling strategy, and construction of semi-structured interview guides. As the research progressed, the CAB provided oversight of emerging insights from data collection, feedback on preliminary themes and theme/codebook refinement, and guidance in the interpretation of findings. Peer debriefing is one of several strategies to address **credibility** and trustworthiness in qualitative research (Bitsch 2005; Guba, 1981; Pitney & Parker, 2009). Our CAB was broader than those typically recognized as peer de-briefers (i.e., academic staff, postgraduate dissertation committees, members of the department, participants in students' seminar presentations; Anney, 2014) because we sought to incorporate insight from mental health professionals and persons with lived experience of mental illness from the design phases through to the dissemination phases of the research.

Sampling

Participants in the current study are a sub-sample of participants from a larger, ongoing study investigating disclosure of serious mental illness in the context of regular employment. Participants in the larger study were recruited through a national health survey (the PULSE), fielded by IBM. The PULSE surveys more than 7,000 U.S. households per month, via phone or internet. The research team added questions to the PULSE to identify and recruit persons who met the following eligibility criteria: working-age (18 to 65 years); diagnosed with bipolar disorder, schizophrenia spectrum disorder, or major depressive disorder; either currently working in a regular job, or had worked in a regular job for six months or longer, post-onset of mental illness and within the prior five years. Those who met the eligibility requirements, and consented to participate, completed a structured phone survey on their experiences in their jobs. At the end of the survey, participants were asked if they were willing to participate in a follow-up semi-structured interview, if eligible. Contact information for those who agreed to further participation was sent to a qualitative team, along with a limited subset of survey response data to determine eligibility for recruitment into the qualitative study.

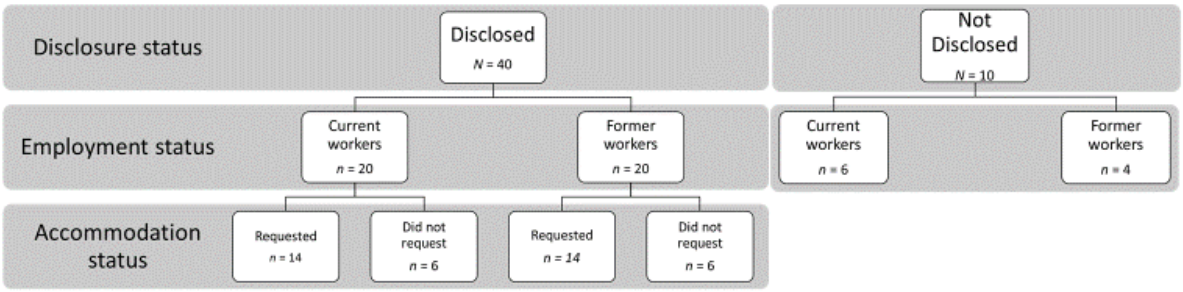
Eligibility for the qualitative study was based on a purposeful sampling frame that stratified the sample by disclosure status (employer knows/does not know about mental illness); employment status (currently employed/previously employed in a regular job); and accommodation status (worker did/did not request job accommodations for their mental illness). This type of sampling is a tool to address the **transferability** component of trustworthiness in qualitative research (Anney, 2014).

Figure 1 in the article displays the sample sizes in each cell of the purposeful sampling framework. To serve as a check to our purposeful sampling frame, we also interviewed a small number of workers who had not disclosed, as shown in Supplemental Figure 1 (N = 10). One of the purposes of these interviews was to address the **credibility** aspect of trustworthiness (Anney, 2014). Specifically, we expected that workers who had not disclosed their SMI to their employers would not be able to share insights on reasons for disclosure or be able to share insights on employer reactions. We wanted to include these

interviews to ensure that data emerging from them did not contradict our prior expectations (Bitsch, 2005). Indeed, when we engaged these workers about why they had not disclosed, and produced qualitative memos about the interviews, we found that they had a homogenous set of reactions that could be tagged as *don't ever tell them* and/or *it's none of their business*.

Based on these preliminary "**negative case**" interviews, we felt that (a) additional interviews were not necessary and (b) the data were not relevant to a study with our specific aims (i.e. (1) describe the contexts in which workers in regular jobs disclose mental illness to their employer; and (2) describe employer responses to disclosure, as perceived by the worker.) Thus, the current study focuses exclusively on the 40 participants whose employers knew about their mental illness. We included an equal number of currently and formerly employed participants, but oversampled those who requested job accommodations. Based on pilot data, we expected greater diversity of employment experiences among those who requested accommodations than among those who did not.

Supplemental Figure 1: Thick Description of Sample (*N* = 40 + 10 negative case interviews)



Data Collection

When we received contact information for potential participants from the firm conducting the larger quantitative survey study, we telephoned potential participants to further screen for eligibility. Workers were eligible to participate in the follow-up interview if their disclosure, employment, and accommodations status had remained stable between the quantitative survey and screening for the follow-up study (usually within a week) and if their sampling stratum had not been filled. The semi-structured

interviews were scheduled and conducted over the phone by two trained interviewers, the PI (first author) and a graduate student. Both interviewers had at least two years of qualitative interviewing experience and at least three years of experience working with individuals with serious mental illness within diverse community contexts (e.g., rural, urban, range of diagnoses, family and individual services). The PI trained in public health and family and human development, and in qualitative, quantitative, and mixed methods. The PI is also the adult child of a parent with a schizophrenia diagnosis who worked in a regular job. The graduate student was a trained master's-level clinician in the final years of a doctorate program in counseling psychology.

The interviewers received more than 40 hours of training for this project, including engagement in mock interviews and conducting 2 pilot interviews. Interviewers were informed of the project's goals and purpose, characteristics of the target population, the importance of professional and clinically sensitive conduct when interviewing participants, and the critical role of interviews in the data collection process. Interviewer training also focused on teaching the following semi-structured interviewing principles: establishing rapport with the participant; obtaining and documenting informed consent; knowing the objectives of the research and interview guides; maintaining strict confidentiality; and managing the interview by neutral probing for detail, listening attentively, and avoiding non-productive discussion. .

Interviews, which were conducted between June 2018 and July 2019, lasted about one hour. Interviews were audio recorded for later transcription. The semi-structured interview guides were designed to (1) build rapport, (2) prompt the workers' disclosure stories, (3) elicit workers' descriptions of their employers' reactions to disclosure, and (4) elicit workers' narratives about the process of requesting and implementing job accommodations. The current study focuses on sections of the interview that prompted disclosure and response narratives. In terms of **dependability** and **confirmability**, we relied upon interview and observational notes written immediately following each interview, as one of several audit trails. Immediately following an interview, the interviewer composed a detailed case study memo describing the interview, including notes about the context of disclosure, employers' responses, job

accommodations, and interview conditions (e.g., interruptions, poor connections). The PI reviewed all memos within a week of an interview completion to monitor quality and progress. To further ensure quality and consistency across the period of data collection, both interviewers listened to and discussed at least one interview weekly.

All interviews were audio-recorded and 39 of 40 interviews were transcribed. (One interview could not be transcribed due to a low-quality recording. For this participant, we analyzed the case study memo.) We followed transcription procedures, based on recommendations described in detail elsewhere (MacQueen & Neidig, 2003). Briefly, there were three people involved in the transcription process: a first transcriber (T1), a second transcriber (T2), and a reconciler (T3). The procedures were as follows: (1) T1 listened to the audio and transcribed directly what was heard into a word processing program; (2) T2, the original interviewer, listened to the audio while reading the transcription from T1 and used track changes to make changes or corrections in the case of discrepancies; (3) T2 and T3 (the other interviewer) met to discuss the discrepancies and come to an agreement on the final transcription.

Analysis and Analytic Plan

Analysis of the data was team-based. As described in the article, our team was composed of researchers with specific interest and experiences with persons with SMI working or trying to work in regular jobs (with distinct positionalities) and persons without any specific interest in the topic, maximizing advantages and disadvantages of each perspective (Hofmann & Barker, 2017). Additionally, throughout the data collection and analysis phases of the project, we relied on team interactions and consultation with the CAB to gain awareness of our preconceptions about the research topic. This process supported our abilities to engage in bracketing, which involved interrogating our own (diverse) assumptions based on our positions and lived experiences vis-à-vis SMI and employment (Hofmann & Barker, 2017). Engagement of multiple investigators with diverse social positions relative to the topic of inquiry is an important form of **triangulation** (Onwuegbuzie & Leech, 2007) to reduce systematic bias and establish **trustworthiness** or **credibility** (Anney, 2014).

As described in the article, we produced two codebooks to facilitate examination of the two aims. The development of each codebook involved weekly meetings of the three authors, who read transcripts and highlighted relevant phrases and ideas based on established observational techniques (Bernard et al., 2017). The highlighted excerpts were sorted into categories, which were clarified and refined through continued observation and discussion. Development of the codebooks also involved regular consultation with the project's CAB, both in the earlier highlighting phase and in the theme refinement phases of the data analysis. This iterative team-based process culminated in labelling and defining key themes and refining the final codebooks.

Once we had a complete draft of a codebook, we proceeded with interrater reliability testing using Dedoose software (Dedoose, 2018). We selected random samples of narrative segments (i.e., complete ideas, usually represented by a participant's talking turn), to be coded by each member of the team. Kappa coefficients were then calculated to assess the level of agreement among coders. The process involved continued team meetings to examine coding discrepancies resulting from the test samples; further refinement of the codebook; and additional interrater reliability tests spaced one to two weeks apart. The process continued until each theme had a kappa coefficient greater than or equal to 0.80. The interrater reliability tests initially were used to aid the process of codebook refinement (Bernard et al.,2017). Additionally, given their temporal spacing and the fact that random samples of texts were comprised of a combination of segments both previously seen and unseen by the coders in earlier stages of the process, the tests (and calculation of kappa coefficients) also serve as code-recode strategy (Anney, 2014), wherein a high degree of similarity between coding sessions and coders increases the dependability aspect of trustworthiness.

Supplemental Results

For step five of the analysis plan, linking themes (Bernard et al., 2017), we focused on the co-occurrence of disclosure and response themes. As an alternative type of theme linking, we could also look at the co-

occurrence of employer response themes (e.g. What proportion of workers who noted compassion in the narrative of their employer's response also noted instrumental help?).

Supplemental Table 1 is a code-by-code matrix showing the co-occurrence of employer response themes within workers' narratives. For example, among the 22 narratives that were coded with the *compassion* theme, 9 (or 40.9%) were also coded with the *instrumentally helpful theme*. The matrix shows that positive employer responses (compassion, experiential understanding instrumental help, respect for confidentiality, and threat-absent) frequently co-occurred with other positive responses, but also sometimes co-occurred with ambiguous (employer emotions, mechanization), or negative responses (lack of respect for confidentiality, threat-present, and stigma).

Supplemental Table 1: Code-by-Code Matrix of Employer Response Themes

Employers' Response(total count)	**1** Compassion	**2** Experiential Understand.	**3** Instrumental help	**4** Respect for confidential.	**5** Threat absent	**6** Employer emotion	**7** Mechaniz.	**8** Lack of respect for confidential.	**9** Threat present	**10** Stigma
1.Compassion (*22)*	\	8 36.4%	9 40.9%	6 27.3%	5 22.7%	5 22.7%	2 9.1%	2 9.1%	4 18.2%	2 9.1%
2. Experiential Understanding (11)	8 72.7%	\	6 54.5%	2 18.2%	1 9.1%	3 27.3%	2 18.2%	1 9.1%	2 18.2%	2 18.2%
3.Instrumental help (12)	9 75.0%	6 50.0%	\	4 33.3%	2 16.7%	2 16.7%	3 25.0%	1 8.3%	1 8.3%	1 8.3%
4.Respect for confidentiality (7)	6 85.7%	2 28.6%	4 57.1%	\	3 42.9%	3 42.9%	1 14.3%	0 0%	0 0%	0 0%
5. Threat Absent (7)	5 71.4%	1 14.3%	2 28.6	3 42.9%	\	2 28.6%	0 0%	0 0%	0 0%	0 0%
6. Employer emotion (11)	5 45.5%	3 27.3%	2 18.2%	3 28.3%	2 18.2%	\	4 36.4%	0 0%	1 9.1%	2 18.2%
7. Mechanization (12)	2 16.7%	2 16.7%	3 35.0%	1 8.3%	0 0%	4 33.3%	\	1 8.3%	2 16.7%	4 33.3%
8. Lack of respect for confidentiality (2)	2 100%	1 50.0%	1 50.0%	0 0%	0 0%	0 0%	1 50%	\	2 100%	1 50.0%
9. Threat-present (5)	4 80.0%	2 40.0%	1 20.0%	0 0%	0 0%	1 20.0%	2 40.0%	2 40.0%	\	3 60.0%
10. Stigma (7)	2 28.6%	2 28.6%	1 14.3%	0 0%	0 0%	2 28.6%	4 57.1%	1 14.3%	3 42.9%	\

Note: Number represent the number of respondents with both themes coded in their narratives. Percentages represent the proportion of respondents in each row with both themes coded.

References

Anney, V. N. (2014). Ensuring the quality of findings of qualitative research Looking at trustworthiness criteria. *Journal of Emerging Trends in Educational Research and Policy Studies*, *5*(2), 272 - 281

Bernard, H. R., Wutich, A., & Ryan, G. W. (2016). *Analyzing qualitative data: Systematic approaches*. SAGE publications

Bitsch, V. (2005). Qualitative research: A grounded theory example and evaluation criteria. Journal of Agribusiness, 23(1), 75-91Guba, 1981

Dedoose. (2018). Dedoose Version 8.0.17, web application for managing, analyzing, and presenting qualitative and mixed method research data. Los Angeles, CA

Hofmann, M., & Barker, C. (2017). On researching a health condition that the researcher has also experienced. *Qualitative psychology*, *4*(2), 139.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277-1288.

Li, D. (2004). Trustworthiness of think-aloud protocols in the study of translation processes. International Journal of Applied Linguistics, 14(3), 301-313. doi: 10.1111/j.1473-4192.2004.00067.x

MacQueen, K. M., McLellan, E., Kay, K., & Milstein, B. (1998). Codebook development for team-based qualitative analysis. *Cam Journal*, *10*(2), 31-36.

Onwuegbuzie, A. J., & Leech, N. L. (2007). Sampling designs in qualitative research: Making the sampling process more public. *Qualitative Report*, *12*(2), 238-254.

Pitney Parker (2009)

Stahl, N. A., & King, J. R. (2020). Expanding approaches for research: Understanding and using trustworthiness in qualitative research. *Journal of Developmental Education*, *44*(1), 26-28.